1. (b)

2. (d)

3. (d)

4. (a)

5. (b)

6. (d)

7. (a)

8. (b)

9. (d)

10. (a)

11. (d)

12. (a)

13. **CLUSTER ANALYSIS CALCULATION**

a). Hierarchical Cluster Analysis

The hierarchical cluster analysis follows three basic steps: 1) calculate the distances, 2) link the clusters, and 3) choose a solution by selecting the right number of clusters.

b). **k-means clustering**

The k-means cluster analysis involves the following 3 steps 1) Randomly select 'c' cluster centres. 2) Calculate the distance between each data point and cluster centers. 3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.

c). **Divisive Clustering**

The divisive clustering algorithm is a top-down clustering approach, initially, all the points in the dataset belong to one cluster and split is performed recursively as one moves down the hierarchy.

14. **CLUSTER QUALITY MEASUREMENT**

If all the data objects in the cluster are highly similar, then the cluster has high quality. We can measure the quality of Clustering by using the Dissimilarity/Similarity metric in most situations. But there are some other methods to measure the Qualities of Good Clustering if the clusters are alike.
1. Dissimilarity/Similarity metric: The similarity between the clusters can be expressed in terms of a distance function, which is represented by d(i, j). Distance functions are different for various data types and data variables. Distance function measure is different for continuous-valued variables, categorical variables, and vector variables. Distance function can be expressed as Euclidean distance, Mahalanobis distance, and Cosine distance for different types of data.
2. **Cluster completeness:** Cluster completeness is the essential parameter for good clustering, if any two data objects are having similar characteristics then they are assigned to the same category of the cluster according to ground truth. Cluster completeness is high if the objects are of the same category.
Let us consider the clustering C1, which contains the sub-clusters s1 and s2, where the members of the s1 and s2 cluster belong to the same category according to ground truth. Let us consider another clustering C2 which is identical to C1 but now s1 and s2 are merged into one cluster. Then, we define

the clustering quality measure, Q, and according to cluster completeness C2, will have more cluster quality compared to the C1 that is, Q(C2, Cg) > Q(C1, Cg).

3. Ragbag: In some situations, there can be a few categories in which the objects of those categories cannot be merged with other objects. Then the quality of those cluster categories is measured by the Rag Bag method. According to the rag bag method, we should put the heterogeneous object into a rag bag category.
Let us consider a clustering C1 and a cluster C ∈ C1 so that all objects in C belong to the same category of cluster C1 except the object o according to ground truth. Consider a clustering C2 which is identical to C1 except that o is assigned to a cluster D which holds the objects of different categories. According to the ground truth, this situation is noisy and the quality of clustering is measured using the rag bag criteria. we define the clustering quality measure, Q, and according to rag bag method criteria   C2, will have more cluster quality compared to the C1 that is, Q(C2, Cg)>Q(C1, Cg).

4. Small cluster preservation: If a small category of clustering is further split into small pieces, then those small pieces of cluster become noise to the entire clustering and thus it becomes difficult to identify that small category from the clustering. The small cluster preservation criterion states that are splitting a small category into pieces is not advisable and it further decreases the quality of clusters as the pieces of clusters are distinctive. Suppose clustering C1 has split into three clusters, C11 = {d1, . . . , dn}, C12 = {dn+1}, and C13 = {dn+2}.
Let clustering C2 also split into three clusters, namely C1 = {d1, . . . , dn−1}, C2 = {dn}, and C3 = {dn+1,dn+2}. As C1 splits the small category of objects and C2 splits the big category which is preferred according to the rule mentioned above the clustering quality measure Q should give a higher score to C2, that is, Q(C2, Cg) > Q(C1, Cg).

15. **CLUSTER ANALYSIS**

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters).

TYPES OF CLUSTER ANALYSIS

Partitioning approach: The partitioning approach constructs various partitions and then evaluates them by some criterion, e.g., minimizing the sum of square errors. It adopts exclusive cluster separation (each object belongs to exactly one group) and uses iterative relocation techniques to improve the partitioning by moving objects from one group to another. It uses a greedy approach and approach at a local optimum. It finds clusters with spherical shapes in small to medium size databases.

Partitioning approach methods:

- k-means
- k-medoids
- CLARINS

2. Density-based approach: This approach is based on connectivity and density functions. It divides the set of objects into multiple exclusive clusters or a hierarchy of clusters. Density-based methods:

- DBSACN
- OPTICS

3. Grid-based approach: This approach quantizes objects into a finite number of cells that form a grid structure. Fast processing time and independent of a number of data objects. Grid-based Clustering method is the efficient approach for spatial data mining problems.
Grid-based approach methods:

- STING
- Wave Cluster
- CLIQUE

4. Hierarchical approach: This creates a hierarchical decomposition of the data objects by using some measures. Hierarchical approach methods:

- Diana
- Agnes
- BIRCH
- CAMELEON